

(W.I.)

Distribution rights by Oxford Molecular Ltd

1

2

11

5

13:sp_vertebrate 14:sp_virus

Job: SCALE 0.521

score distribution.

LOBULIN-1 (FRAGMENT). 5.3

0

1

Q5	199	2.2	1/1 11	VOL110	QUOTIENTIAL REFERENCE	2.000	1.0

S0 SEQUENCE 525 AA; 60798 MW; 4890EACE CRC32; 000

3

Matches: 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0

→

DD I M V I S N S P E I V A L I F S L L B S F A B C B S V S A I G K A V I E N D F K V U I E V E G N K R C E S E A W E E N K Z E 00

0Y 1 MWISKSPFIVLIFSLLLSCVSAVGRKQYERDPROQYEOCORCESEATEEREQE 60

61 QGEGRCEREXKEBQKQDEBELQRYDDEUGKCEUEQDQGGKEDQDCEQKCKWEQYKEJERGE 120

0V 61 0CE0RCEREYKE00RO0EEEL0R0Y00C0GRCE0E000G0RE000C0BK0WE0YK0E0ERGE 120

D5 121 HENYHHKKNRSEEEGGQRRNNPYFFPKRRSFQTRFDEEGNFKILQRF AENS PPLKGIN 180

(TM)

Distribution rights by Oxford Molecular Ltd

t generated

~~(1-525) FROM US0~~

154

Gap 11

131253

Listing first 45

a-issued

Mean 33.

ved by analysis of the total score distribution

2.9

ALIGNMENTS

45 96

XXXXXX

Sequencer

Seque

GENE

AF

II

88

20

INFO

SEQUEN

Jerry Mat

atches

13

13-

610

2—

•

—

WORLDWIDE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database: MARCH, using Smith-Waterman algorithm
Run on: Sat May 13 11:45:14 2000; Maspar time 58.52 Seconds
130.618 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-8
Description: (1-599) from US09331631:pep (1 of 4)
Perfect Score: 4332
Sequence: 1 MVRKRSACVLLFSLFSLFSG.....FNSNPQESFVSRRORORASE 590

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfillset

Statistics: Mean 34.226; Variance 174.925; scale 0.196

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	4119	95.1	587	1	US-07-955-Sequence 23, Applicati	0.00e+00
2	2101	48.5	566	1	US-07-955-Sequence 22, Applicati	7.38e-162
3	2101	48.5	566	1	US-07-955-Sequence 2, Applicati	7.38e-162
4	968	22.3	605	1	US-07-955-Sequence 24, Applicati	1.71e-67
5	832	19.2	571	1	US-07-955-Sequence 25, Applicati	2.62e-56
6	776	17.9	410	1	US-07-955-Sequence 26, Applicati	1.01e-51
7	493	11.4	421	1	US-07-955-Sequence 27, Applicati	8.72e-29
8	223	5.1	1162	2	US-08-728-Sequence 2, Applicati	8.99e-08
9	166	3.8	1898	2	US-08-800-Sequence 94, Applicati	1.23e-03
10	166	3.8	1898	2	US-08-056-Sequence 94, Applicati	1.23e-03
11	162	3.7	788	2	US-08-918-Sequence 4, Applicati	2.33e-03
12	138	3.2	186	2	US-08-557-Sequence 43, Applicati	1.07e-01
13	127	2.9	361	1	US-08-415-Sequence 4, Applicati	5.84e-01
14	123	2.8	737	1	US-08-185-Sequence 4, Applicati	1.07e+00
15	123	2.8	737	1	US-08-185-Sequence 2, Applicati	1.07e+00
16	119	2.7	434	1	US-08-337-Sequence 3, Applicati	1.96e+00
17	119	2.7	2237	2	US-08-223-Sequence 48, Applicati	1.96e+00
18	119	2.7	2237	2	US-08-455-Sequence 47, Applicati	1.96e+00
19	119	2.7	2339	2	US-08-455-Sequence 47, Applicati	1.96e+00
20	119	2.7	2339	2	US-08-223-Sequence 47, Applicati	1.96e+00
21	112	2.6	357	1	US-08-910-Sequence 2, Applicati	5.57e+00
22	112	2.6	357	1	US-08-552-Sequence 2, Applicati	5.57e+00
23	112	2.6	357	3	PCT-US95-0 Sequence 2, Applicati	5.57e+00

RESULT ID	1	US-07-955-905A-23	STANDARD:	PRT:	587 AA.
24	112	2.6	369	2	US-08-474-Sequence 32, Applicati
25	112	2.6	369	1	US-07-688-Sequence 32, Applicati
26	112	2.6	369	3	PCT-US91-0 Sequence 31, Applicati
27	112	2.6	914	1	US-08-484-Sequence 2, Applicati
28	112	2.6	914	1	US-08-484-Sequence 42, Applicati
29	112	2.6	3144	2	US-08-457-Sequence 6, Applicati
30	112	2.6	3144	1	US-08-246-Sequence 6, Applicati
31	112	2.6	3144	1	US-08-453-Sequence 9, Applicati
32	109	2.5	156	3	PCT-US95-0 Sequence 9, Applicati
33	109	2.5	156	1	US-08-552-Sequence 9, Applicati
34	109	2.5	156	1	US-08-552-Sequence 73, Applicati
35	110	2.5	240	3	PCT-US95-0 Sequence 73, Applicati
36	109	2.5	303	2	US-08-459-Sequence 2, Applicati
37	109	2.5	303	2	US-08-460-Sequence 2, Applicati
38	109	2.5	303	1	US-08-109-Sequence 5, Applicati
39	109	2.5	303	1	US-08-185-Sequence 15, Applicati
40	107	2.5	356	1	US-08-552-Sequence 15, Applicati
41	109	2.5	356	1	US-08-910-Sequence 15, Applicati
42	109	2.5	356	2	US-08-910-Sequence 2, Applicati
43	109	2.5	387	1	US-08-539-Sequence 2, Applicati
44	107	2.5	387	1	US-08-539-Sequence 2, Applicati
45	107	2.5	387	1	US-08-539-Sequence 2, Applicati

ALIGNMENTS

Sequence 23, Application US/07955905A

Sequence 23, Application US/07955905A

Patent No. 5770433

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND

TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND

NUMBER OF SEQUENCES: 28

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/955,905A

FILING DATE: 21-JAN-1993

CLASSIFICATION: 435

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 587 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORGANISM: Gossypium hirsutum

FEATURE:

NAME/KEY: Protein

LOCATION: 1..587

OTHER INFORMATION: /note="Vicilin from G. hirsutum"

SEQUENCE 587 AA: 69497 MW: 1625753 CN:

Query Match 95.1%; Score 4119; DB 1; Length 587;

Best Local Similarity 95.1%; Pred. No. 0.00e+00;

Matches 561; Conservative 17; Mismatches 9; Indels 3; Gaps 2;

Db 1 MVRKRSACVLLFSLFSLFSLCSANDPGRKDDDPKRYEDCRRRGEWDTGKQKQ 60
1 MVRKRSACVLLFSLFSLFSLCSAKDFPGRGDDDPKRYEDCRRRCEWDTGKQKQ 60

1	3654	96.0	566	2	522477	vicillin precursor - c	0.00e+00
2	1919	50.4	509	2	506359	alpha-globulin type A	0.00e+00
3	1910	50.2	565	2	508058	alpha-globulin type B	0.00e+00
4	1853	48.7	588	1	FMGNAB	alpha-globulin B prec	0.00e+00
5	881	23.1	573	2	FA32334	globulin-1S, GLB1S -	3.82e-14
6	848	22.3	582	2	B53234	vicillin-like storage	1.82e-13
7	830	21.8	433	2	S00567	vicillin precursor (cl	7.85e-13
8	827	21.7	439	1	FMSCB1	beta-conglycinin beta	3.17e-13
9	826	21.7	459	2	S08505	vicillin - garden pea	5.04e-13
10	810	21.3	605	1	FMSYBA	beta-conglycinin alph	8.49e-13
11	805	21.1	605	2	S20007	beta-conglycinin alph	8.64e-13
12	776	20.4	571	2	S00566	convicillin precursor	5.92e-12
13	771	20.2	463	2	A27288	vicillin precursor - f	5.99e-12
14	771	20.2	463	2	S06309	vicillin precursor (cl	5.99e-12
15	764	20.1	438	2	S375757	vicillin, 47K - garden	1.53e-12
16	732	19.2	410	1	FMFMB	vicillin B precursor -	3.98e-11
17	724	19.0	448	1	S4756	vicillin-like storage	1.59e-11
18	702	18.4	524	2	J017130	6XK sucrose-binding p	3.95e-10
19	697	18.3	483	2	T06459	6XK sucrose-binding p	3.93e-10
20	682	17.5	450	2	S15675	globulin-2 precursor	3.84e-10
21	644	16.9	445	2	S00281	canavalin - sword bea	1.37e-97
22	643	16.9	445	2	J02264	canavalin - jack bean	2.16e-97
23	634	16.6	639	2	B24810	beta-conglycinin alph	1.31e-95

24	593	15.6	637	2	S55221	globoxin	beg1 pcd	1.65e-87	
25	531	13.9	196	2	S02281	convicillin	(clone) prec	2.33e-75	
26	551	13.9	236	2	T01662	globoxin-1	- maize (f	1.07e-64	
27	470	12.3	436	2	A23498	phasellon	type alpha	1.53e-63	
28	466	12.2	407	2	T02258	globoxin	- maize	8.98e-63	
29	435	11.4	275	1	FMPWA	vicillin A precursor	-	7.75e-57	
30	434	11.4	421	1	A24810	phasellon	beta chain	1.20e-56	
31	426	11.2	414	2	S60190	vicillin -	zambia fertu	4.01e-55	
32	418	11.0	430	2	S10156	alpha-phasellon	precu	1.33e-53	
33	391	10.3	540	2	S21825	vicillin-like	storage	1.67e-48	
34	365	9.6	296	2	T06372	convicillin	precursor	1.23e-43	
35	35	9.5	240	2	T06399	convicillin	precursor	3.33e-28	
36	230	6.0	128	1	F16334	beta-conglycinin	alph	4.98e-18	
37	232	5.8	124	1	FMPW4	vicillin, 14k componen		1.48e-18	
38	195	5.1	613	2	S27770	hypochemical protein		1.31e-13	
39	151	5.0	136	2	T02870	globoxin	2 precursor	5.81e-13	
40	168	4.9	184	1	FEFB	phasellon	- kidney be	1.76e-12	
41	183	4.8	1038	1	T02634	rep	protein	homolog	1.11e-11
42	170	4.5	395	2	S35565	sex-determining	prote	-	1.21e-09
43	170	4.5	395	2	S43344	sex-determining	prote	-	1.21e-09
44	173	4.5	708	2	A53185	G-box-binding	factor	-	4.13e-10
45	171	4.5	1094	2	S59313	protein kinase	- blm		8.45e-10

ALIGNMENTS

ENTRY	1
TITLE	
ORGANISM	
DATE	
ACCESSIONS	
REFERENCE	
#authors	
#journal	
#title	
S22477	#type complete
vicilin precursor - cacao	
vicilin precursor - cacao	#common_name cacao
30-Sep-1993	#sequence_revision 30-Sep-1993
17-Mar-1999	#text_change
S22477; S22478; S18105; S22050	
S22477	
McHenry, L.; Fritze, P.J.	
Plant Mol. Biol. (1992) 18:1173-1176	
Comparison of the structure and nucleotide sequence of	
vicilin genes of cocoa and cotton raise questions about	
vicilin evolution.	

```

#accession      S22477
#molecule_type DNA
#residues       1-566  ##label MCH
#cross-references EMBL:X62625
#accession      S22478

```

GENETICS

CLASSIFICATION *superfamily glycinin

KEYWORDS seed; storage protein

```

1-24      #domain signal sequence #status predicted #label SICV
25-566    #product; vicillin #status predicted #label MAT
SUMMARY   #length 566 #molecular-weight 65513 #checksum 3693

```

Query Match	96.0%;	Score 3654;	DB 2;	length 566;
Best Local Similarity	96.7%;	Pred. No. 0.00e+00;		
Matches	504;	Conservative	10;	Mismatches 6; Indels 1; Gaps 1

Db	1	MYISKSPFYILVILFSLLSFALLCSGSAVYGRKQYERPRDQYEECCQRCSEAEATEERDE	60
	1	MYISKSPFYILVILFSLLSFALLCSGSAVYGRKQYERPRDQYEECCQRCSEAEATEERDE	60
Qy	1	MYISKSPFYILVILFSLLSFALLCSGSAVYGRKQYERPRDQYEECCQRCSEAEATEERDE	60
Db	61	QCQGRCEKREKEDQROOEELQRYOQCQGRCEQOQGRQOQOQCKKCKWEQYQGEORGE	120
	61	QCQGRCEKREKEDQROOEELQRYOQCQGRCEQOQGRQOQOQCKKCKWEQYQGEORGE	120
Qy	61	QCQGRCEKREKEDQROOEELQRYOQCQGRCEQOQGRQOQOQCKKCKWEQYQGEORGE	120
Db	121	HEHYHNHKKRRSEEEEGQQRNNPYFPFKRRSFQTRFDEEGNFILQRFANSPPLKGIN	180
	121	HEHYHNHKKRRSEEEEGQQRNNPYFPFKRRSFQTRFDEEGNFILQRFANSPPLKGIN	180
Qy	121	HEHYHNHKKRRSEEEEGQQRNNPYFPFKRRSFQTRFDEEGNFILQRFANSPPLKGIN	180

Db 181 DYRLAMFANPTFLPHHCOAEIYVINGKTIITVTHEKESYNORGTWVSPAGS 240
181 DYRLAMFANPTFLPHHCOAEIYVINGKTIITVTHEKESYNORGTWVSPAGS 240
Db 241 TVYVVSQDNOEKLTIYALVLPVNSPGKYLEFFPAGNNKPESSYGAFFVETVNTORE 300
241 TVYVVSQDNOEKLTIYALVLPVNSPGKYLEFFPAGNNKPESSYGAFFVETVNTORE 300
Db 301 KLEELLEDORGGKROGGOGGFRKAKPEOIRAIISOATSPRRHGERLAINLLSSPYVS 360
301 KLEELLEDORGGKROGGOGGFRKAKPEOIRAIISOATSPRRHGERLAINLLSSPYVS 360
Db 361 NONGHFEACPEDESOFOFNDVAVSAFKLNOGAIFVPHYSKATFVVTGOGYAOAC 420
361 NONGHFEACPEDESOFOFNDVAVSAFKLNOGAIFVPHYSKATFVVTGOGYAOAC 420
Db 421 PHLSSQSGSGSGRODRREDEESEEETFGFEQVAKPLSPDVAVAGHATVFFASKD 480
421 PHLSSQSGSGSGRODRREDEESEEETFGFEQVAKPLSPDVAVAGHATVFFASKD 480
Db 481 OPLNAVAGLNAONNORIFLAGKRLVROMDSEAKELSGV 521
481 OPLNAVAGLNAONNORIFLAGKRLVROMDSEAKELSGV 521
OY 481 OPLNAVAGLNAONNORIFLAGKRLVROMDSEAKELSGV 521
RESULT 2
ENTRY 506398 #type complete
TITLE alpha-globulin type A precursor - upland cotton
ALTERNATE_NAMES seed storage protein
ORGANISM #formal_name Gossypium hirsutum #common_name upland cotton
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
ACCESSIONS 506398
REFERENCE 506398
#authors Chien, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
#journal Plant Mol. Biol. (1987) 9:533-546
#title Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Sequences and genomic organization of the alpha globulin (vicilin) genes of cottonseed.
#accession 506398
#status not compared with conceptual translation
#molecule_type DNA
#residues 1-605 #label CHL
CLASSIFICATION #superfamily glycylin
FEATURE 1-24
25-605
#domain signal sequence #status predicted #label SIG
#product alpha-globulin type A #status predicted #label MAT
SUMMARY #length 605 #molecular_weight 71049 #checksum 4376
Query Match 50.44: Score 1919; DB 2; Length 605;
Best Local Similarity 57.0%; Pred. No. 0.00e+00;
Matches 269; Conservative 101; Mismatches 87; Indels 15; Gaps 13;
Db 79 DPOARRQDCRHOQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECQRCQOWOE 133
79 DPOARRQDCRHOQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECQRCQOWOE 133
OY 37 DPOQYEQOCORCESEATEEREEOCEORCEREYKQOQEEELQROYOQOCGRQO--EQ 95
37 DPOQYEQOCORCESEATEEREEOCEORCEREYKQOQEEELQROYOQOCGRQO--EQ 95
Db 134 OPPER-KQOCVRECRYOYEDPMKGERENKMRSEEESEDEQOQRNPPYFH--RRSQ 191
134 OPPER-KQOCVRECRYOYEDPMKGERENKMRSEEESEDEQOQRNPPYFH--RRSQ 191
OY 96 OCOGRQOQOCCKKCEQYKEOE--RGEHEN-YHNHKNRSEEEQOQRNPPYFPRKRSQ 153
96 OCOGRQOQOCCKKCEQYKEOE--RGEHEN-YHNHKNRSEEEQOQRNPPYFPRKRSQ 153
Db 192 EFRREHGNFVLOFADKHLRGINFERIALILEANPTFLPHHCOAEIYVINGK 251
192 EFRREHGNFVLOFADKHLRGINFERIALILEANPTFLPHHCOAEIYVINGK 251
OY 154 TFRDEEGNFKILOFPAESPLKGINDRMLAFENPTFLPHHCOAEIYVINGK 213
154 TFRDEEGNFKILOFPAESPLKGINDRMLAFENPTFLPHHCOAEIYVINGK 213
Db 252 TVTFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 311
252 TVTFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 311
OY 252 TVTFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 311
252 TVTFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 311
Db 321 TITFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 273
321 TITFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 273
OY 321 TITFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 273
321 TITFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 273
Db 370 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 370
370 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 370
OY 370 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 370
370 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 370
Db 333 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 333
333 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 333
OY 333 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 333
333 RQENRQSLRIFRREILEAVNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 333

Db 371 SQGATSPHKGSEGTAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 430
371 SQGATSPHKGSEGTAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 430
OY 334 SQGATSPHKGSEGTAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 392
334 SQGATSPHKGSEGTAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 392
Db 431 SIPIPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 488
431 SIPIPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 488
OY 393 AIPIPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 452
393 AIPIPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 452
Db 489 KRYAOLSTGNLFFVAPGHPYTFVASONEDLGLGFLYNGQDKRIFVACK 540
489 KRYAOLSTGNLFFVAPGHPYTFVASONEDLGLGFLYNGQDKRIFVACK 540
OY 453 QVYKAPLSPGDVAVAPAGHATVFFASKDQPLNAVAFGL--MNQNNRITLAOR 503
453 QVYKAPLSPGDVAVAPAGHATVFFASKDQPLNAVAFGL--MNQNNRITLAOR 503
RESULT 3
ENTRY 508059 #type fragment
TITLE alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
ALTERNATE_NAMES seed storage protein
ORGANISM #formal_name Gossypium hirsutum #common_name upland cotton
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
ACCESSIONS 508059
REFERENCE 508059
#authors Chien, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
#journal Plant Mol. Biol. (1987) 9:533-546
#title Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Sequences and genomic organization of the alpha globulin (vicilin) genes of cottonseed.
#accession 508059
#status not compared with conceptual translation
#molecule_type DNA
#residues 1-509 #label CHL
CLASSIFICATION #superfamily glycylin
SUMMARY #length 509 #checksum 9476
Query Match 50.28: Score 1910; DB 2; Length 509;
Best Local Similarity 56.9%; Pred. No. 0.00e+00;
Matches 261; Conservative 95; Mismatches 92; Indels 11; Gaps 10;
Db 1 DPOARRQDCRHOQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECQRCQOWOE 54
1 DPOARRQDCRHOQOE--ERLRPHCEQSCREY--EKQOQOQPD--KQFKECQRCQOWOE 54
OY 37 DPOQYEQOCORCESEATEEREEOCEORCEREYKQOQEEELQROYOQOCGRQO--EQ 96
37 DPOQYEQOCORCESEATEEREEOCEORCEREYKQOQEEELQROYOQOCGRQO--EQ 96
Db 55 OPPER-KQOCVRECRYOYEDPMKGERENKMRSEEESEDEQOQRNPPYFH--RRSQ 111
55 OPPER-KQOCVRECRYOYEDPMKGERENKMRSEEESEDEQOQRNPPYFH--RRSQ 111
OY 97 OCOGRQOQOCCKKCEQYKEOE--RGEHEN-YHNHKNRSEEEQOQRNPPYFPRKRSQ 155
97 OCOGRQOQOCCKKCEQYKEOE--RGEHEN-YHNHKNRSEEEQOQRNPPYFPRKRSQ 155
Db 112 EFRREHGNFVLOFADKHLRGINFERIALILEANPTFLPHHCOAEIYVINGK 171
112 EFRREHGNFVLOFADKHLRGINFERIALILEANPTFLPHHCOAEIYVINGK 171
OY 156 TFRDEEGNFKILOFPAESPLKGINDRMLAFENPTFLPHHCOAEIYVINGK 215
156 TFRDEEGNFKILOFPAESPLKGINDRMLAFENPTFLPHHCOAEIYVINGK 215
Db 172 TVTFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 231
172 TVTFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 231
OY 216 TITFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 275
216 TITFTTHEKESYNVPGVYVIRPAGSTVYLANODNREKLTIYALVLPVNSPGKYLEFF 275
Db 232 SORPOSTYARFRELLEPFAFNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 290
232 SORPOSTYARFRELLEPFAFNTRSEQDELPGRQSRROQO--GMFRKASQOIRAL 290
OY 276 NKPESYGAFFVETVNTOREKLEELLEDORGGKROGGOGGFRKAKPEOIRAI 335
276 NKPESYGAFFVETVNTOREKLEELLEDORGGKROGGOGGFRKAKPEOIRAI 335
Db 291 EATSPREKSGERFAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 350
291 EATSPREKSGERFAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 350
OY 336 SQGATSPHKGSEGTAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 395
336 SQGATSPHKGSEGTAFNLLSOTPRYSNONGRFEACPRNFQOOLREVDSVVAFEIUNG 395
Db 351 VPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 410
351 VPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 410
OY 396 AIPIPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 455
396 AIPIPHYSKATFVVTGOGYAOACPHLSROSSDW--SRREE--EOEEVEARRSOY 455
Db 411 RSRLSRGDIFFVAPGHPYTFVASONEDLGLGFLYNGQDKRIFVACK 449
411 RSRLSRGDIFFVAPGHPYTFVASONEDLGLGFLYNGQDKRIFVACK 449
OY 456 KRYAOLSTGNLFFVAPGHPYTFVASONEDLGLGFLYNGQDKRIFVACK 503
456 KRYAOLSTGNLFFVAPGHPYTFVASONEDLGLGFLYNGQDKRIFVACK 503

(TM)

Distribution rights by Oxford Molecular, Inc.

Sat May 13 14:41:21 2000

14.761 Million cell updates/sec

3)

PAM 150

131233 seqs, 1233004, 10020000

Listing first 45 summaries

1: 5A COMB 2: 5B COMB 3: PCT_COMB 4: backfill

Mean 23.260; Variance 99.417, scale 0.000

and is derived by analysis of the total score distribution"

1
2
3
4
5

4:

ALIGNMENTS

88

388

88

38

8

28

333

88

333

Que

...

3

XY
YD

XX

MUSE RELEASE (™)

Release 3.1A John F. Collins, Biocomputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MUSE_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 10:45:58 2000; MasPar time 7.09 Seconds
Tabular output not generated. 398,931 Million cell updates/sec

Title: >US-09-331-631-7
Description: 09-1401 from US09331631.pep (3 of 3)
Perfect Score: 471
Sequence: 1 LOROXOCCGRCOECOCQOQOR.....HENYHNKKRSEEGQOR 60

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 32.973; Variance 71.563; scale 0.461

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	471	100.0	566	2 S22477	vicillin precursor - c	8.50e-68
2	188	39.9	605	2 S06398	alpha-globulin type A	8.81e-17
3	181	38.4	509	2 S08059	alpha-globulin type B	1.30e-15
4	168	35.7	588	1 FWCNAB	alpha-globulin B prec	1.84e-13
5	125	26.5	1038	2 T02634	rep protein homolog -	1.20e-06
6	121	25.7	613	2 S27770	hypothetical protein	4.82e-05
7	116	24.6	564	2 S37241	legumin B - fava bean	2.69e-05
8	114	24.2	495	1 I37062	involucrin S - gortill	5.32e-05
9	114	24.2	605	1 I37061	involucrin M - gortill	5.32e-05
10	114	24.2	635	1 I37060	involucrin L - gortill	5.32e-05
11	114	24.2	835	1 I57441	involucrin - oranga	5.32e-05
12	111	23.6	522	1 I37037	involucrin - common g	1.46e-04
13	111	23.6	551	2 S51941	prunin 1 precursor -	1.46e-04
14	111	23.6	708	2 AS3185	G-box-binding factor	1.46e-04
15	110	23.4	1390	2 T14004	trif protein - slime	2.05e-04
16	109	23.1	1094	2 T44933	protein kinase - slim	2.86e-04
17	107	22.7	1457	2 T14577	protein kinase YAKA (5.57e-04
18	105	22.3	285	1 I46207	involucrin - dog	1.08e-03
19	105	22.3	905	1 R6BYS5	regulatory protein SN	1.50e-03
20	104	22.1	119	2 A28892	Mopa box protein - mo	1.50e-03
21	104	22.1	1655	2 T13998	gene mastermind prote	1.50e-03
22	103	21.9	289	2 S74632	hypothetical protein	2.08e-03
23	103	21.9	395	2 S43344	sex-determining prote	2.08e-03

24	103	21.9	395	2 S35655	sex-determining prote	2.08e-03
25	101	21.4	1154	2 S69206	regulator protein vhl	3.98e-03
26	100	21.2	47	2 JC5557	arginine/glutamate-r1	5.51e-03
27	100	21.2	255	2 A50637	mezoioite antigen LPM	5.51e-03
28	100	21.2	637	2 S35221	globulin Bgl1 precus	5.51e-03
29	100	21.2	1898	1 A45973	trichohyalin - human	5.51e-03
30	100	21.2	2150	2 S71629	sensory transduction	5.51e-03
31	99	21.0	758	2 S54522	hypothetical protein	7.60e-03
32	99	21.0	930	2 T08588	hypothetical protein	7.60e-03
33	99	21.0	1761	2 T13675	hypothetical protein	7.60e-03
34	98	20.8	203	2 S09559	legumin A (clone PC04	1.05e-02
35	98	20.8	507	2 T06452	probable legumin A pr	1.05e-02
36	98	20.8	517	1 FMPMLA	legumin A precursor -	1.05e-02
37	98	20.8	560	1 A40125	involucrin - pygmy ch	1.05e-02
38	98	20.8	1010	2 T13167	Loia-like protein - f	1.05e-02
39	97	20.6	544	1 I36911	involucrin L - dourou	1.44e-02
40	97	20.6	625	2 A34615	profilaggrin - rat (f	1.44e-02
41	97	20.6	2359	2 T03094	A-kinase anchor prote	1.44e-02
42	96	20.4	398	2 S21883	DNA-binding protein P	1.98e-02
43	96	20.4	496	2 S33791	ARS-binding protein -	1.98e-02
44	96	20.4	1021	2 T10748	mannan endo-1,4-beta	1.98e-02
45	95	20.2	338	2 S04321	legumin B (clone PC03	2.72e-02

ALIGNMENTS

RESULT 1
ENTRY S22477 #type complete
TITLE vicillin precursor - cacao
ORGANISM #formal name Theobroma cacao #common name cacao
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999

ACCESSIONS S22477: S22478: S18105: S22050
REFERENCE S22477
#authors Mchenry, L.; Fritz, P.J.
#journal Plant Mol. Biol. (1992) 18:1173-1176
#title Comparison of the structure and nucleotide sequence of
vicillin genes of cocoa and cotton raise questions about
vicillin evolution.

#cross-references MIMD:92284309
#accession S22477
#molecule_type DNA
#residues 1-566 #label MCH
#cross-references EMBL:X62625

#accession S22478
#molecule_type mRNA
#residues 1-452 #label MC2
#cross-references EMBL:X62626

GENETICS
#introns 211/1: 269/3: 296/3: 391/3: 502/1
CLASSIFICATION #superfamily glycinin
KEYWORDS seed; storage protein
FEATURE
1-24
25-566

SUMMARY
#domain signal sequence #status predicted #label SIC\
#product vicillin #status predicted #label MAT
#length 566 #molecular-weight 65513 #checksum 3693

Query Match 100.0%: Score 471. DB 2: Length 566;
Best Local Similarity 100.0%: Pred. No. 8.50e-68;
Matches 60; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

Db 81 LOROXOCCGRCOECOCQOQORCKCMEQYKEQERGHENYHNKKRSEEGQOR 140
Qy 81 LOROXOCCGRCOECOCQOQORCKCMEQYKEQERGHENYHNKKRSEEGQOR 140

RESULT 2
ENTRY S06398 #type complete
TITLE alpha-globulin type A precursor - upland cotton
ALTERNATE_NAMES seed storage protein
ORGANISM #formal name Gossypium hirsutum #common name upland cotton
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993

WUJ May 13 09:24:00 2000

Release 3.1A John F. Collins, Biocomputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 10:41:27 2000; Maspar time 239.76 Seconds
17.351 Million cell updates/sec

Tabular output not generated.

Title: <05-09-331-631-7
Description: 471 481-1401 from 0509331631 pep (3 of 3)
Perfect Score: 1 LOROYOCGRCGCEQOQGOR.....HENYHNKKNRSEEGGOR 60
Sequence: 1 LOROYOCGRCGCEQOQGOR.....HENYHNKKNRSEEGGOR 60

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database:

sptrembl12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mmc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 32.878; Variance 66.854; scale 0.492

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	471	100.0	525	10	Q43358	VICILIN PRECURSOR.	2.19e-72
2	133	28.2	411	5	P91419	CODED FOR BY C. ELEGAN	1.92e-08
3	125	26.5	1038	5	O60983	REP-LIKE.	3.90e-07
4	119	25.3	171	11	O61118	GLUTAMINE REPEAT PROTE	3.59e-06
5	116	25.3	870	10	O92W13	PV100.	1.08e-05
6	116	24.6	564	10	Q43673	LEGUMIN, LEGUMIN-RELAT	1.08e-05
7	116	24.6	905	5	O18260	ZC101.1 PROTEIN.	1.08e-05
8	114	24.2	495	6	Q28425	INVOLUCRIN.	2.22e-05
9	114	24.2	605	6	Q28424	INVOLUCRIN.	2.22e-05
10	114	24.2	635	6	Q28423	INVOLUCRIN.	2.22e-05
11	112	23.8	838	5	O910C9	RAS INTERACTING PROTEI	4.57e-05
12	111	23.6	551	10	Q43607	PRONIN PRECURSOR.	6.54e-05
13	110	23.4	304	4	O15409	CAGH44 (FRAGMENT).	9.36e-05
14	110	23.4	388	5	O16500	CO3A7.4 PROTEIN.	9.36e-05
15	110	23.4	388	5	O16501	CO3A7.7 PROTEIN.	9.36e-05
16	110	23.4	1380	5	O77033	TRFA.	1.34e-04
17	109	23.1	1094	5	Q23915	PROTEIN KINASE.	2.72e-04
18	107	22.7	1457	5	O44011	PROTEIN KINASE YAKA.	7.81e-04
19	104	22.1	139	11	O62006	OPA REPEAT (FRAGMENT).	7.81e-04
20	104	22.1	406	2	O87306	ELPB2.	7.81e-04

21	104	22.1	445	5	O16511	CO3A7.14 PROTEIN.	7.81e-04
22	104	22.1	632	4	O94918	VEIN PROTEIN PRECURSOR	7.81e-04
23	104	22.1	652	4	O15410	CAGH45.	7.81e-04
24	104	22.1	1655	5	O24754	MASTERMIND.	7.81e-04
25	104	22.1	2023	4	O75557	OPA-CONTAINING PROTEIN.	7.81e-04
26	104	22.1	2074	11	O88542	OPA-CONTAINING PROTEIN	7.81e-04
27	103	21.9	289	2	P72769	HYPOTHETICAL 31.2 KD P	1.11e-03
28	103	21.9	420	11	O35861	SEX DETERMINING PROTEI	1.11e-03
29	103	21.9	438	5	O16502	CO3A7.8 PROTEIN.	1.11e-03
30	102	21.7	57	4	O914J1	SCA1 PROTEIN (FRAGMENT	1.57e-03
31	102	21.7	218	2	O44782	ERPB.	1.57e-03
32	102	21.7	378	2	O08264	ERPB2.	1.57e-03
33	102	21.7	388	5	O31328	ERPB.	1.57e-03
34	102	21.7	416	11	O35957	SEX DETERMINING PROTEI	1.57e-03
35	102	21.7	418	11	O35862	SEX DETERMINING PROTEI	1.57e-03
36	102	21.7	418	11	O35863	SEX DETERMINING PROTEI	1.57e-03
37	101	21.4	624	5	O94486	ORF DGI040 (FRAGMENT).	2.22e-03
38	101	21.4	624	5	O94486	ORF DGI040 (FRAGMENT).	2.22e-03
39	101	21.4	2253	11	O9WU43	SILENCING MEDIATOR OF	2.22e-03
40	101	21.4	2462	11	O9WU43	SILENCING MEDIATOR OF	2.22e-03
41	101	21.4	2473	11	O9WU42	AT MOTIF-BINDING FACTO	2.22e-03
42	101	21.4	3726	11	O61329	SEX DETERMINING PROTEI	3.13e-03
43	100	21.2	234	11	O35859	EMBRYO GLOBULIN.	3.13e-03
44	100	21.2	637	10	O03678	EMBRYO GLOBULIN.	3.13e-03
45	100	21.2	2150	5	O23863	HISTIDINE KINASE A.	3.13e-03

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	Q43358	100.0%	Best Local	525	Q43358	VICILIN PRECURSOR.	2.19e-72	
AC	Q43358	100.0%	Matches	60	Q43358	VICILIN PRECURSOR.	2.19e-72	
DT	01-NOV-1996	100.0%			01-NOV-1996	(TREMBLrel. 01, last sequence update)		
DT	01-NOV-1996	100.0%			01-NOV-1996	(TREMBLrel. 01, last sequence update)		
DE	01-NOV-1999	100.0%			01-NOV-1999	(TREMBLrel. 12, last annotation update)		
DE	VICILIN	100.0%			VICILIN	PRECURSOR.		
GN	CSV	100.0%			CSV			
OS	Theobroma cacao (Cacao).	100.0%			Theobroma cacao (Cacao).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	100.0%			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphylliphytes; Spermatophytes; Magnoliophyta; eudicotyledons;	100.0%			euphylliphytes; Spermatophytes; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eustosids II; Malvales; Malvaceae; Theobroma.	100.0%			core eudicots; Rosidae; eustosids II; Malvales; Malvaceae; Theobroma.			
RN	[1]	100.0%			[1]			
RP	SEQUENCE FROM N.A.	100.0%			SEQUENCE FROM N.A.			
RC	TISSUE-LEAVES:	100.0%			TISSUE-LEAVES:			
RX	MEDLINE: 92286309.	100.0%			MEDLINE: 92286309.			
RA	McHENRY L., FRITZ P.J.;	100.0%			McHENRY L., FRITZ P.J.;			
RT	"Comparison of the structure and nucleotide sequences of vicilin genes	100.0%			"Comparison of the structure and nucleotide sequences of vicilin genes			
RT	of cocoa and cotton raise questions about vicilin evolution."	100.0%			of cocoa and cotton raise questions about vicilin evolution."			
RL	Plant Mol. Biol. 18:1173-1176(1992).	100.0%			Plant Mol. Biol. 18:1173-1176(1992).			
DR	EMBL: X62625; CAA44493.1;	100.0%			EMBL: X62625; CAA44493.1;			
DR	EMBL: X62626; CAA44494.1;	100.0%			EMBL: X62626; CAA44494.1;			
DR	HSSP: P02853; 2PHL.	100.0%			HSSP: P02853; 2PHL.			
DR	MENDEL: 30919; Thecc:1188;30919.	100.0%			MENDEL: 30919; Thecc:1188;30919.			
DR	PFAW: PF00546; seedstore_7s; 1.	100.0%			PFAW: PF00546; seedstore_7s; 1.			
KT	Signal.	100.0%			Signal.			
FT	SIGNAL.	100.0%			SIGNAL.			
FT	CHAIN	100.0%			CHAIN			
SO	SEQUENCE	100.0%			SEQUENCE			
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		
DT	01-MAY-1997	100.0%			01-MAY-1997	(TREMBLrel. 03, last sequence update)		

RELEASE

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_p protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 12:12:57 2000: Maspar time 59.01 Seconds
7.246 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-23
Description: 287
Perfect Score: 1 RSGRCECRQCRHRECPWETQECMRRCRRG 33
Sequence: 1

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT_COMB 4:backfiles1

Statistics: Mean 22.066; Variance 101.322; scale 0.218

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	287	100.0	34	2	US-08-822-Sequence 1, Applicatio	3.78e+17
2	87	30.3	188	2	US-08-569-Sequence 11, Applicatio	3.85e+00
3	87	30.3	188	1	US-08-469-Sequence 11, Applicatio	3.85e+00
4	87	30.3	188	2	US-08-609-Sequence 11, Applicatio	3.85e+00
5	86	30.0	566	1	US-07-955-Sequence 2, Applicatio	4.59e+00
6	86	30.0	566	1	US-07-955-Sequence 22, Applicatio	1.56e+01
7	79	27.5	587	1	US-07-955-Sequence 23, Applicatio	3.67e+01
8	74	25.8	55	2	US-08-609-Sequence 3, Applicatio	3.67e+01
9	74	25.8	55	2	US-08-469-Sequence 3, Applicatio	3.67e+01
10	74	25.8	55	1	US-08-469-Sequence 3, Applicatio	3.67e+01
11	74	25.8	188	2	US-08-569-Sequence 5, Applicatio	3.67e+01
12	74	25.8	188	2	US-08-469-Sequence 5, Applicatio	3.67e+01
13	74	25.8	188	1	US-08-469-Sequence 7, Applicatio	3.67e+01
14	74	25.8	188	2	US-08-609-Sequence 7, Applicatio	3.67e+01
15	74	25.8	195	1	US-08-469-Sequence 7, Applicatio	3.67e+01
16	74	25.8	195	1	US-08-569-Sequence 7, Applicatio	3.67e+01
17	71	24.7	3025	4	5223423-3Patent No. 5223423	6.09e+01
18	70	24.4	77	2	US-08-726-Sequence 17, Applicatio	7.20e+01
19	70	24.4	384	2	US-08-673-Sequence 11, Applicatio	7.20e+01
20	70	24.4	384	2	US-08-614-Sequence 11, Applicatio	7.20e+01
21	70	24.4	642	2	US-08-926-Sequence 25, Applicatio	7.20e+01
22	70	24.4	642	1	US-08-414-Sequence 25, Applicatio	7.20e+01
23	69	24.0	900	2	US-08-813-Sequence 4, Applicatio	8.51e+01

24	69	24.0	1101	2	US-08-469-Sequence 96, Applicatio	8.51e+01
25	69	24.0	1118	1	US-08-278-Sequence 2, Applicatio	8.51e+01
26	69	24.0	1118	2	US-08-838-Sequence 2, Applicatio	8.51e+01
27	69	24.0	1118	1	US-07-934-Sequence 2, Applicatio	8.51e+01
28	69	24.0	1122	2	US-08-838-Sequence 6, Applicatio	8.51e+01
29	69	24.0	1122	1	US-08-218-Sequence 6, Applicatio	8.51e+01
30	68	23.7	37	2	US-08-505-Sequence 7, Applicatio	1.01e+02
31	68	23.7	37	2	US-09-018-Sequence 7, Applicatio	1.01e+02
32	68	23.7	301	2	US-08-656-Sequence 25, Applicatio	1.01e+02
33	67	23.3	158	2	US-08-670-Sequence 6, Applicatio	1.01e+02
34	66	23.0	488	1	US-08-477-Sequence 1, Applicatio	1.40e+02
35	66	23.0	488	1	US-08-484-Sequence 1, Applicatio	1.40e+02
36	66	23.0	524	1	US-08-477-Sequence 2, Applicatio	1.40e+02
37	66	23.0	524	1	US-08-484-Sequence 2, Applicatio	1.40e+02
38	66	23.0	524	1	US-08-464-Sequence 2, Applicatio	1.40e+02
39	66	23.0	524	1	US-08-243-Sequence 3, Applicatio	1.40e+02
40	66	23.0	670	1	US-08-477-Sequence 3, Applicatio	1.40e+02
41	66	23.0	670	1	US-08-477-Sequence 3, Applicatio	1.40e+02
42	66	23.0	670	1	US-08-484-Sequence 4, Applicatio	1.40e+02
43	66	23.0	769	1	US-08-484-Sequence 4, Applicatio	1.40e+02
44	66	23.0	769	1	US-08-484-Sequence 4, Applicatio	1.40e+02
45	66	23.0	769	1	US-08-243-Sequence 4, Applicatio	1.40e+02

ALIGNMENTS

RESULT 1 STANDARD: PRT: 34 AA.

ID US-08-822-561-1

AC xxxxxx

XX

DT

XX Sequence 1, Application US/08822561

DE

XX Sequence 1, Application US/08822561

CC Patent No. 590187

CC GENERAL INFORMATION:

CC APPLICANT: Duwick, John P.

CC APPLICANT: Rood, Tracy A.

CC APPLICANT: Rao, Aravula G.

CC TITLE OF INVENTION: ANTIMICROBIAL PEPTIDES AND PLANT DISEASE

CC TITLE OF INVENTION: RESISTANCE BASED THEREON

CC NUMBER OF SEQUENCES: 3

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.

CC STREET: 700 Capital Square, 400 Locust Street

CC CITY: Des Moines

CC STATE: Iowa

CC COUNTRY: USA

CC ZIP: 50309

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.30

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/822,561

CC FILING DATE: 19-MAR-1997

CC CLASSIFICATION: 800

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: US 08/442,910

CC FILING DATE: 17-MAY-1995

CC APPLICATION NUMBER: US 07/536,127

CC FILING DATE: 05-JUN-1990

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Yates, Michael J.

CC REGISTRATION NUMBER: 36,063

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (515) 248-4800

CC TELEFAX: (515) 248-4844

CC INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
 CC LENGTH: 34 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: linear
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 34 AA: 4188 MW: 3921 CN:

Query Match 100.0%; Score 287; DB 2; Length 34;
 Best Local Similarity 100.0%; Pred. No. 3,78e-17;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 RSGRGCRRCCLRRHGGOPWETOECMRRCRRG 33
 1 RSGRGCRRCCLRRHGGOPWETOECMRRCRRG 33

RESULT 2
 ID US-08-569-063C-11 STANDARD: PRT: 188 AA.
 AC xxxxxx
 DT
 DE Sequence 11, Application US/08569063C
 CC Sequence 11, Application US/08569063C
 CC Patent No. 5928939
 CC GENERAL INFORMATION:
 CC APPLICANT: ERIKSSON, Ulf
 CC APPLICANT: OLOFSSON, Birgitta
 CC APPLICANT: ALITALO, Kari
 CC APPLICANT: PAJUSOLA, Katri
 CC TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
 CC TITLE OF INVENTION: DNA CODING THEREFOR
 CC NUMBER OF SEQUENCES: 23
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.
 CC STREET: 1200 G Street, N.W., Suite 700
 CC CITY: Washington
 CC STATE: DC
 CC COUNTRY: USA
 CC ZIP: 20005
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: floppy disk
 CC OPERATING SYSTEM: IBM PC compatible
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/569,063C
 CC FILING DATE: 06-DEC-1995
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/469,427
 CC FILING DATE: 06-JUN-1995
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/397,651
 CC FILING DATE: 01-MAR-1995
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: EVANS, Joseph D
 CC REGISTRATION NUMBER: 26,269
 CC REFERENCE/DOCKET NUMBER: 1064/41979CP3
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (202) 628-8800
 CC TELEFAX: (202) 628-8844
 CC INFORMATION FOR SEQ ID NO: 11:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 188 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC HYPOTHETICAL: NO
 CC ORIGINAL SOURCE:
 CC SIGNAL SOURCE:

TISSUE TYPE: human fibrosarcoma
 CC SEQUENCE 188 AA: 21261 MW: 161002 CN:

Query Match 30.3%; Score 87; DB 2; Length 188;
 Best Local Similarity 40.6%; Pred. No. 3,85e+00;
 Matches 13; Conservative 7; Mismatches 9; Indels 3; Gaps 3;

DB 137 SPRPLCPR-CTOHQ-RP-DPRTCRCRCRRS 165
 2 SGRGCRRCCLRRHGGOPWETOECMRRCRRG 33

RESULT 3
 ID US-08-469-427A-11 STANDARD: PRT: 188 AA.
 AC xxxxxx
 DT
 DE Sequence 11, Application US/08469427A
 CC Sequence 11, Application US/08469427A
 CC Patent No. 5607918
 CC GENERAL INFORMATION:
 CC APPLICANT: ERIKSSON, Ulf
 CC APPLICANT: OLOFSSON, Birgitta
 CC APPLICANT: ALITALO, Kari
 CC APPLICANT: PAJUSOLA, Katri
 CC TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND
 CC TITLE OF INVENTION: DNA CODING THEREFOR
 CC NUMBER OF SEQUENCES: 17
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Evenson, McKeown, Edwards & Lenahan
 CC STREET: 1200 G Street, N.W., Suite 700
 CC CITY: Washington
 CC STATE: DC
 CC ZIP: 20005
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: floppy disk
 CC OPERATING SYSTEM: IBM PC compatible
 CC SOFTWARE: Patentin Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/469,427A
 CC FILING DATE: 06-JUN-1995
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/397,651
 CC FILING DATE: 01-MAR-1995
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: EVANS, Joseph D
 CC REGISTRATION NUMBER: 26,269
 CC REFERENCE/DOCKET NUMBER: 41979CP2
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (202) 628-8800
 CC TELEFAX: (202) 628-8844
 CC INFORMATION FOR SEQ ID NO: 11:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 188 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC HYPOTHETICAL: NO
 CC ORIGINAL SOURCE:
 CC TISSUE TYPE: human fibrosarcoma
 CC SEQUENCE 188 AA: 21261 MW: 161002 CN:

Query Match 30.3%; Score 87; DB 1; Length 188;
 Best Local Similarity 40.6%; Pred. No. 3,85e+00;
 Matches 13; Conservative 7; Mismatches 9; Indels 3; Gaps 3;

DB 137 SPRPLCPR-CTOHQ-RP-DPRTCRCRCRRS 165

WIRE (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSParch_bp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 09:00:19 2000; Maspar time 78.64 Seconds
8.907 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-25
Description: (1-23) FROM US09331631.pep
Perfect Score: 177
Sequence: 1 MMRARFPLLLGLVFLASVSF 23

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 08
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 33.887; Variance 71.622; scale 0.473
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

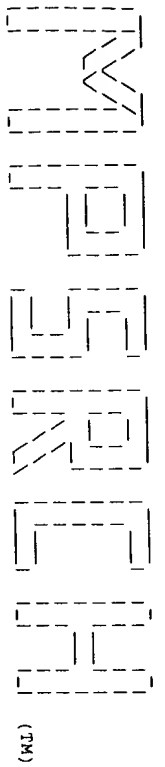
SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	177	100.0	605	1	GLCA_SOYBN	BETA-CONGLYCININ, ALPH	2.46e+15
2	170	96.0	639	1	GLCX_SOYBN	BETA-CONGLYCININ, ALPH	3.71e+14
3	138	78.0	421	1	PHSB_PHAVU	PHASEOLIN, BETA-TYPE P	6.39e+09
4	138	78.0	436	1	PHSA_PHAVU	PHASEOLIN, ALPHA-TYPE	6.39e+09
5	136	76.8	439	1	GLCB_SOYBN	BETA-CONGLYCININ, BETA	1.33e+08
6	120	67.8	571	1	CVCB_PEA	CONVICILIN PRECURSOR	4.08e+06
7	114	64.4	386	1	CVCB_PEA	CONVICILIN PRECURSOR	3.28e+05
8	111	62.7	445	1	CANA_CANLN	CANAVALIN PRECURSOR	9.18e+05
9	111	62.7	445	1	CANA_CANLN	CANAVALIN PRECURSOR	9.18e+05
10	105	59.3	459	1	VCLC_PEA	VICILIN PRECURSOR	6.96e+04
11	94	53.1	463	1	VCLC_PEA	VICILIN PRECURSOR	2.52e+02
12	84	47.5	547	1	CP78_MAIZE	CYTROCHROME P450 78A1	5.57e+01
13	82	46.3	659	1	VST2_HEVBU	STRUCTURAL PROTEIN 2 P	1.36e+00
14	81	45.8	502	1	V752_BORBU	HYPOTHELICAL PROTEIN B	1.36e+00
15	81	45.8	614	1	AH12_ARAHY	ALLERGEN ARA H 1, CLON	1.36e+00
16	81	45.8	624	1	AH12_ARAHY	ALLERGEN ARA H 1, CLON	1.36e+00
17	81	45.8	909	1	LDL1_XENLA	LOW-DENSITY LIPOPROTEI	1.83e+00
18	80	45.2	141	1	LSA_BPP2	PROTEIN LISA	1.83e+00
19	80	45.2	449	1	SCRC_RAT	SECRETIN RECEPTOR PREC	2.44e+00
20	79	44.6	67	1	YPHE_BACSU	HYPOTHELICAL 7.5 KD PR	2.44e+00
21	79	44.6	405	1	WCAD_ECOLI	PUTATIVE COLANIC ACID	4.35e+00
22	77	43.5	124	1	VBF3_YEAST	VERY HYPOTHELICAL 13.2	4.35e+00
23	77	43.5	660	1	VST2_HEVBU	STRUCTURAL PROTEIN 2 P	4.35e+00

SEQ ID NO: 25

RESULT	ID	Accession	Standard	PRT	605 AA	ALIGNMENTS
24	77	43.5	660	1	VST2_HEVBU	STRUCTURAL PROTEIN 2 P
25	77	43.5	660	1	VST2_HEVBU	STRUCTURAL PROTEIN 2 P
26	76	42.9	213	1	GLP1_PHANI	IMMUNOGLOBULIN-RELATED
27	76	42.9	214	1	GLP1_PHANI	IMMUNOGLOBULIN-RELATED
28	76	42.9	336	1	MRAT_CHLTR	GERMIN-LIKE PROTEIN PR
29	75	42.4	769	1	ITB2_BOVIN	PROSPERO-N-ACETYLIMBRO
30	75	42.4	979	1	ITB2_BOVIN	CELL SURFACE ADHESION
31	74	41.8	173	1	AMD_MOUSE	PEPTIDYL-GLYCINE ALPHA
32	74	41.8	412	1	GLUP_BRUB	PATHOGENESIS-RELATED P
33	74	41.8	603	1	GLUP_BRUB	GLUCOSE/GALACTOSE TRAN
34	74	41.8	106	1	NU5M_GORGO	NADH-UBIQUINONE OXIDOR
35	73	41.2	202	1	PSIF_ECOLI	PHOSPHATE STARVATION-I
36	73	41.2	393	1	YBBD_BACSU	HYPOTHELICAL 42.7 KD L
37	73	41.2	642	1	YBBD_BACSU	HYPOTHELICAL 42.7 KD L
38	73	41.2	837	1	YBBD_BACSU	HYPOTHELICAL 42.7 KD L
39	73	41.2	892	1	YBBD_BACSU	HYPOTHELICAL 42.7 KD L
40	73	41.2	904	1	TORS_ECOLI	ENDO-1,4-BETA-XYLANASE
41	73	41.2	3033	1	POLG_HCVJ6	GENOME PROTEIN TORS (E
42	72	40.7	142	1	LCA_CAVPO	ALPHA-LACTALBUMIN PREC
43	72	40.7	405	1	LPPB_HAEN	OUTER MEMBRANE ANTIGEN
44	72	40.7	669	1	FP51_YEAST	GLYCEROL UPTAKE/EFFLUX
45	72	40.7	690	1	PERO_DROME	PEROXIDASE PRECURSOR

1
ID GLCA_SOYBN STANDARD; PRT; 605 AA.
AC P13916;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BETA-CONGLYCININ, ALPHA CHAIN PRECURSOR.
OS Glycine max (soybean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC eudicotyledons; Spermatophytes; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Glycine.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COTYLEDON;
RA MEDLINE; 91353860.
RX Sebastiani F.L., Farrel L.B., Schuler M.A., Beachy R.N.;
RT "Complete sequence of a cDNA of alpha subunit of soybean beta-
conglycinin".
RT Plant Mol. Biol. 15:197-201(1990).
CC - FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC - SUBUNIT: THE ALPHA-, ALPHA-, AND BETA-SUBUNTS ASSOCIATE IN
CC VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC - SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND
CC VACUOLAR PROTEIN BODIES.
CC - SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.slb.ch).
CC EMBL: X17698; CAA35691.1;
CC PIR: S14681; FMSYBA.
DR HSPF: P50477; ICAH.
DR PFM: P00546; Seedstore-7s; 1.
KW Seed storage protein; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 22
FT PROPEP 23 62
FT CHAIN 63 605 BETA-CONGLYCININ, ALPHA CHAIN.
FT CARBOHYD 261 261 POTENTIAL.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:51:56 2000; Maspar time 6.17 Seconds
252.411 Million cell updates/sec

Tabular output not generated.

Title: >US-09-331-631-23
Description: (1-33) from 450331631.pep
Perfect Score: 287
Sequence: 1 RSGRGECRCROCLRRHGGPWTQECMRCRRRG 33

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 30.542; Variance 60.292; Scale 0.507

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	287	100.0	33	2	A41822	antimicrobial peptide	7.27e-39
2	92	32.1	442	2	H72266	astB/chur-related pro	5.27e-03
3	91	31.7	122	2	C53234	globulin-10 - maize	7.50e-03
4	91	31.7	582	2	B53234	vicillin-like storage	7.50e-03
5	91	31.7	588	1	FWCMB	alpha-globulin B prec	2.14e-02
6	88	30.7	509	2	S08059	alpha-globulin type B	4.27e-02
7	86	30.0	221	2	C34768	ORF2 protein - Ort vi	4.27e-02
8	86	30.0	566	2	S22477	vicillin precursor - C	6.02e-02
9	85	29.6	236	2	T01662	globulin-1 - maize (f	6.02e-02
10	85	29.6	407	2	T02258	globulin-1 - maize	6.02e-02
11	85	29.6	540	2	S21825	vicillin-like storage	6.02e-02
12	85	29.6	573	2	A53234	globulin-1S, GBS1S -	8.46e-02
13	84	29.3	52	2	I50722	neuron-glia adhesion	2.33e-01
14	82	28.6	527	2	A54985	BMP receptor precursor	1.66e-01
15	81	28.2	358	2	I50506	gene wnt8b protein -	2.33e-01
16	81	28.2	428	2	I51680	xmt-8b - African cla	3.24e-01
17	80	27.9	112	2	I51578	gene wnt-8b protein -	3.24e-01
18	80	27.9	525	2	A56316	acu-8 protein - Neuro	4.52e-01
19	79	27.5	91	2	S21673	protamine 2 - pig	4.52e-01
20	79	27.5	344	2	E70950	probable transposase	6.28e-01
21	78	27.2	48	2	S68952	omega-hordothionin -	6.28e-01
22	78	27.2	82	2	S55779	gamma-thionin precurs	8.70e-01
23	77	26.8	45	2	B58208	protamine II-1 - pain	8.70e-01

24	77	26.8	58	2	A58208	protamine II-1 - pain	8.70e-01
25	77	26.8	187	2	S64323	ribosomal protein S4	8.70e-01
26	77	26.8	605	2	S06398	alpha-globulin type A	8.70e-01
27	77	26.8	4976	2	T14155	peptide synthetase ho	1.20e+00
28	76	26.5	45	2	B69810	hypothetical protein	1.20e+00
29	76	26.5	47	2	E58208	protamine II-4 - pain	1.20e+00
30	76	26.5	51	2	B75130	lsu ribosomal protein	1.20e+00
31	76	26.5	51	2	C71001	probable ribosomal pr	1.20e+00
32	76	26.5	102	2	S33332	protamine P2 - pygmy	1.20e+00
33	76	26.5	102	2	S33333	protamine P2 - gotill	1.20e+00
34	76	26.5	102	2	S33331	protamine P2 - chimpa	1.20e+00
35	74	25.8	45	2	C58208	protamine II-2 - pain	2.29e+00
36	74	25.8	47	2	F58208	protamine II-5 - pain	2.29e+00
37	74	25.8	102	1	HS0UP2	sperm histone P2 prec	2.29e+00
38	74	25.8	188	2	JC4680	vascular endothelial	2.29e+00
39	74	25.8	1073	2	S56220	probable membrane pro	2.29e+00
40	74	25.8	1306	2	T13592	hypothetical protein	3.14e+00
41	73	25.4	58	2	S10755	protamine St2b - hors	3.14e+00
42	73	25.4	62	2	S10754	protamine St2a - hors	3.14e+00
43	73	25.4	92	2	S13132	protamine 2 precursor	3.14e+00
44	73	25.4	102	2	S33336	protamine P2 - thesus	3.14e+00
45	73	25.4	4006	2	T09070	probable tenascin X -	3.14e+00

ALIGNMENTS

RESULT 1
ENTRY A41822 *type complete
TITLE antimicrobial peptide - maize
ORGANISM Zea mays
DATE 28-May-1992 *sequence-revision 28-May-1992 *text-change 30-Sep-1993

ACCESSIONS
REFERENCE A41822
#authors DAVICK, J.P., ROOD, T., ROO, A.G., MARSHAK, D.R.
#submission submitted to the Protein Sequence Database, May 1992
#accession A41822

##status preliminary
##molecule_type protein
##residues 1-33 ##label DUV
SUMMARY ##length 33 ##molecular_weight 4131 ##checksum 3118

Query Match 100.0%; Score 287; DB 2; Length 33;
Best Local Similarity 100.0%; Pred. No. 7.27e-39;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 RSGRGECRCROCLRRHGGPWTQECMRCRRRG 33
QY 1 RSGRGECRCROCLRRHGGPWTQECMRCRRRG 33

RESULT 2
ENTRY H72266 *type complete
TITLE astB/chur-related protein - Thermotoga maritima (strain MSB8)
ORGANISM Thermotoga maritima
DATE 11-Jun-1999 *sequence-revision 11-Jun-1999 *text-change 11-Jun-1999

ACCESSIONS
REFERENCE H72266
#authors NELSON, K.E.; CLAYTON, R.A.; GILL, S.R.; GWINN, M.L.; DODSON, J.R.; HAFT, D.H.; HICKEY, E.K.; PETERSON, J.D.; NELSON, W.C.; KETCHUM, K.A.; McDONALD, L.; UTERBACK, T.R.; MALEK, J.A.; LINHER, K.D.; GARRETT, M.M.; STEWART, A.M.; COTTON, M.D.; PRATT, M.S.; PHILLIPS, C.A.; RICHARDSON, D.; HEIDELBERG, J.; SUTTON, G.G.; FLEISCHMANN, R.D.; WHITE, O.; SALZBERG, S.L.; SMITH, H.O.; VENTER, J.C.; FRASER, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
#accession H72266
#status preliminary
##cross-references PubMed:99287316
##molecule_type DNA

WISREX

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:35:40 2000; Maspar time 73.66 seconds
24.806 Million cell updates/sec
Tabular output not generated

Title: >US-09-331-631-21
Description: 132-911 from US09331631.pep
Perfect Score: 453
Sequence: 1 TENPCAQRCLQSCQDEPDDL.....DGGATNQRHPGERTGRQP 60

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 33.775; Variance 51.458; scale 0.656
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	453	100.0	614	1 AH11_ARAHY	ALLERGEN ARA H 1, CLON	1.78e-92
2	396	87.4	626	1 AH12_ARAHY	ALLERGEN ARA H 1, CLON	1.84e-77
3	120	26.5	639	1 GLCX_SOYBN	BETA-CONGLYCININ, ALPH	1.15e-09
4	119	26.3	605	1 GLCX_SOYBN	BETA-CONGLYCININ, ALPH	1.86e-09
5	93	20.5	680	1 KALM_HUMAN	KALLMANN SYNDROME PROT	2.88e-04
6	85	18.8	676	1 KALM_HUMAN	KALLMANN SYNDROME PROT	8.57e-03
7	77	17.0	4753	1 LRP_GAEEL	LOW-DENSITY LIPOPROTEI	2.14e-01
8	76	16.8	605	1 VCLA_GOSHI	VICILIN GC72-A PRECURS	3.16e-01
9	75	16.6	128	1 RNP_CHOHO	RIBONUCLEASE PANCREATI	4.64e-01
10	75	16.6	211	1 Y121_BURCE	INSERTION ELEMENT IS40	4.64e-01
11	75	16.6	649	1 V70K_EPMV	70 KD PROTEIN,	4.64e-01
12	74	16.3	147	1 IP21_SOLTU	PROTEINASE INHIBITOR T	6.80e-01
13	74	16.3	229	1 VHEL_PVMR	PROBABLE HELICASE (ORF	6.80e-01
14	74	16.3	320	1 YBIB_ECOLI	HYPOTHETICAL 35.0 KD P	6.80e-01
15	74	16.3	488	1 CRUA_BRANA	CRUCIFERIN PRECURSOR (6.80e-01
16	74	16.3	588	1 VCLB_GOSHI	VICILIN C72 PRECURSOR	6.80e-01
17	74	16.3	1106	1 STC_DROME	SHUTTLE CRAFT PROTEIN.	6.80e-01
18	73	16.1	477	1 ETV1_MOUSE	ERB1 PROTEIN (ETS TRAN	9.92e-01
19	72	15.9	753	1 YJ05_GAEEL	HYPOTHETICAL 85.1 KD N	1.44e+00
20	72	15.9	1955	1 AGRI_CHICK	AGRIIN PRECURSOR.	1.44e+00
21	71	15.7	154	1 IP27_SOLTU	PROTEINASE INHIBITOR T	2.09e+00
22	71	15.7	204	1 IP22_CAPAN	WOOND-INDUCED PROTEINA	2.09e+00
23	71	15.7	607	1 YN92_YEAST	POTATIVE TRANSCRIPTION	2.09e+00

SCAIDN021

ID	1	STANDARD	PRT	614 AA.
AC	P4237			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	ALLERGEN ARA H 1, CLONE P17 (ARA H I).			
OS	Arachis hypogaea (Peanut).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;			
OC	Arachis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CV, FLORINER;			
RX	MEDLINE; 96013631.			
RA	Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannan G.A.;			
RT	"Recombinant peanut allergen Ara h I expression and IgE binding in			
RT	patients with peanut hypersensitivity.";			
RL	J Clin Invest. 96:1115-1121(1995).			
CC	- SIMILARITY: TO 75 SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,			
CC	CONVICILIN, CONGLYCININ, ETC.).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb.ch/announce/			
CC	or send an email to license@isb.slb.ch).			
CC	-----			
CC	EMBL; L38853; AA60336.1; .			
DR	HSSP; P50477; ICAM.			
DR	PFAM; PF00546; Seedstore_7s.1.			
KW	Allergen.			
SQ	SEQUENCE 614 AA: 70283 MW: 10DACF217ECC5F31 CRC64;			

Query Match 100.0%; Score 453; DB 1; Length 614;
Best Local Similarity 100.0%; Pred. No. 1.78e-92;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 32 TENPCAQRCLQSCQDEPDDLKAKACSSRCTKLEYDPRCVYDTGATNQRHPGERTGRQP 91
|||||
32 TENPCAQRCLQSCQDEPDDLKAKACSSRCTKLEYDPRCVYDTGATNQRHPGERTGRQP 91

RESULT 2

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

24 80 45.2 44.9 2 516319 secretin receptor - r 7.49e+00
25 79 44.6 67 2 B69936 hypothalamic protein 9.70e+00
26 79 44.6 402 2 T13499 NADH dehydrogenase su 9.70e+00
27 79 44.6 405 2 G64971 hypothalamic protein 1.25e+01
28 78 44.1 492 2 T02458 hypothalamic protein 1.25e+01
29 78 44.1 684 2 T12151 NADH dehydrogenase - 1.25e+01
30 78 44.1 684 2 T13491 NADH dehydrogenase - 1.25e+01
31 78 44.1 685 2 T12129 NADH dehydrogenase - 1.25e+01
32 78 44.1 704 2 T13503 NADH dehydrogenase F 1.25e+01
33 78 44.1 732 2 T14233 NADH dehydrogenase - 1.25e+01
34 78 44.1 734 2 T12602 NADH dehydrogenase - 1.25e+01
35 78 44.1 734 2 T13785 NADH dehydrogenase - 1.25e+01
36 77 43.5 124 2 S45788 probable membrane pro 1.62e+01
37 77 43.5 246 2 E64627 hypothalamic protein 1.62e+01
38 77 43.5 271 2 T11692 NADH dehydrogenase su 1.62e+01
39 77 43.5 433 2 G70345 hypothalamic protein 1.62e+01
40 77 43.5 680 2 VHMW2 structural protein 1.62e+01
41 77 43.5 686 2 T13680 NADH dehydrogenase - 1.62e+01
42 76 42.9 336 2 D71474 probable muramoyl-pen 2.09e+01
43 76 42.9 381 2 T13701 NADH dehydrogenase su 2.09e+01
44 76 42.9 534 1 S75101 hypothetical protein 2.09e+01
45 76 42.9 696 2 T13767 NADH dehydrogenase - 2.09e+01

ALIGNMENTS

RESULT 1
ENTRY FWSYBA #type complete
TITLE beta-conglycinin alpha chain precursor - soybean
ORGANISM #formal_name Glycine max #common_name soybean
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999

ACCESSIONS S14681: S74124; S06714
REFERENCE S14681
#authors Sebastiani, F.L.; Farrell, L.B.; Schuler, M.A.; Beachy, R.N.
#journal Plant Mol. Biol. (1990) 15:197-201
#title Complete sequence of a cDNA of alpha subunit of soybean beta-conglycinin.

#cross-references MUID:91353680
#accession S14681
#molecule_type mRNA

#residues 1-605 #label SEB
#molecule_type mRNA

REFERENCE S74123
#authors Shutoy, A.D.; Kakhovskaya, I.A.; Bastrygina, A.S.; Bulmaga, V.P.; Horstmann, C.; Muentz, K.

#journal Eur. J. Biochem. (1996) 241:221-228
#title Limited proteolysis of beta-conglycinin and glycinin, the 7S and 11S storage globulins from soybean [Glycine max (L.) Merr.]. Structural and evolutionary implications.

#cross-references MUID:97054613
#accession S74124
#molecule_type protein

#residues 189-196, 'H', 196, 'N', 200, 'X', 202-203, 397-408, 'X', 410, 'X', 412-417, 'X', 419-420, 'X', 484-501, 'Y', 503 #label SHU
#experimental_source seed
#suprafamily glycinin
#glycoprotein: seed; storage protein

FEATURES
1-22 #domain signal sequence #status predicted #label SIG
23-62 #domain propeptide #status predicted #label PRO
63-605 #product beta-conglycinin alpha chain #status predicted #label MAT

261-517 #binding site carbohydrate (Asn) (covalent) #status predicted
SUMMARY #length 605 #molecular_weight 70293 #checksum 6280

Query Match 100.0% Score 177: DB 1: Length 605:
Best Local Similarity 100.0% Pred. No. 5.92e-13:
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MMRAFPLLGLVFLASVSF 23
|||||

SEQ ID NO: 25

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	177	100.0	605	1	FWSYBA beta-conglycinin alph 5.92e-13	
2	177	100.0	605	2	S20007 beta-conglycinin alph 5.92e-13	
3	170	96.0	639	2	B24810 beta-conglycinin alph 6.41e-12	
4	138	78.0	421	2	A24810 phaseolin beta chain 2.56e-07	
5	138	78.0	430	2	S10156 alpha-phaseolin precu 2.56e-07	
6	138	78.0	436	2	A23498 phaseolin type alpha 4.87e-07	
7	136	76.8	439	1	FWSYCB beta-conglycinin beta 7.55e-05	
8	120	67.8	571	2	S00566 convicillin precursor 4.75e-04	
9	114	64.4	296	2	T06572 convicillin precursor 4.75e-04	
10	111	62.7	445	2	S00281 canavalin - sword bea 1.18e-03	
11	111	62.7	445	2	JO2264 canavalin - jack bean 1.18e-03	
12	105	59.3	459	2	S08505 vicilin - garden pea 7.03e-03	
13	94	53.1	463	2	A27288 vicilin precursor - f 1.68e-01	
14	94	53.1	463	2	S06309 vicilin precursor (cl 2.61e-01	
15	84	47.5	547	1	S1475 cytochrome P450 cyp8 3.41e+00	
16	83	46.9	438	2	S35757 vicilin, 47k - garden 4.44e+00	
17	82	46.3	400	2	G72350 sugar ABC transporter 4.44e+00	
18	82	46.3	400	2	F72350 structural protein 2 4.44e+00	
19	82	46.3	659	1	B42132 hypothetical protein 2 5.77e+00	
20	81	45.8	502	2	G70193 hypothetical protein 5.77e+00	
21	81	45.8	909	1	ORXLL1 LysA receptor 1 precu 7.49e+00	
22	80	45.2	141	2	E55855 LysA protein - phage 7.49e+00	
23	80	45.2	293	2	H72064 metal transport prote 7.49e+00	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

WIDEOR (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:59:23 2000; Maspar time 6.82 Seconds
380,618 Million cell updates/sec

Tabular output not generated.
Title: >US-09-331-631-24
Description: 425
Perfect Score: 1 DEDDRRGHSLQOCVORCRQ.....DDQOGRHHEEEOGRGRG 55

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues
Post-processing: Minimum Match 08
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 32.380; Variance 69.376; scale 0.467

Prod. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	425	100.0	637	2	S35221	globulin Bgl1 precurs	2,456-60
2	174	40.9	132	2	C53234	globulin-10 - maize	6,976-15
3	174	40.9	582	2	B53234	vicilin-like storage	6,976-15
4	170	40.0	236	2	T01662	globulin-1 - maize (f	3,266-14
5	168	39.5	407	2	T02258	globulin-1 - maize	7,046-14
6	168	39.5	540	2	S21825	vicilin-like storage	7,046-14
7	135	31.8	588	2	A53234	globulin-1S, GUB1S -	7,046-14
8	135	31.8	588	2	FWCNA8	alpha-globulin B prec	1,666-08
9	131	30.8	509	2	S08059	alpha-globulin type A	7,066-08
10	107	25.2	605	2	S06388	alpha-globulin type B	3,186-04
11	105	24.7	957	2	D70835	hypothetical glycine-	6,236-04
12	101	23.8	1549	1	A40691	trichohyalin - sheep	2,376-03
13	92	21.6	1021	1	T10748	mannan endo-1,4-beta-	4,406-02
14	92	21.6	1023	1	S12519	glutactin - fruit fly	4,406-02
15	92	21.6	1898	1	A45973	trichohyalin - human	1,136-01
16	89	20.9	566	2	S22477	vicilin precursor - c	1,136-01
17	89	20.9	613	2	S27770	hypothetical protein -	1,136-01
18	89	20.9	625	2	A34615	profilaggrin - rat (f	1,136-01
19	88	20.7	1200	2	T15921	hypothetical protein	1,556-01
20	88	20.7	2248	1	D42088	adenylate cyclase (EC	2,126-01
21	87	20.5	330	2	S28102	rlx protein - Staphyl	3,926-01
22	85	20.0	450	2	G01158	tyrosine kinase activ	3,926-01
23	85	20.0	678	2	S54308	DNA binding protein -	3,926-01

SEQ ID NO: 24

RESULT	ENTRY	DESCRIPTION	SEQUENCE	REVISION	TEXT CHANGE
24	85	20.0	788	2	I51530
25	84	19.8	325	2	JH0189
26	84	19.8	380	2	S36789
27	84	19.8	382	2	A33640
28	84	19.8	524	2	JQ1730
29	84	19.8	551	2	S51941
30	84	19.8	1255	1	A24571
31	83	19.5	558	2	S32951
32	83	19.5	559	2	S52941
33	83	19.5	559	2	S52953
34	83	19.5	559	2	S52954
35	83	19.5	559	2	S52955
36	83	19.5	559	2	S52956
37	83	19.5	559	2	S52957
38	83	19.5	559	2	S52958
39	83	19.5	559	2	S52959
40	83	19.5	559	2	S52960
41	83	19.5	559	2	S52961
42	83	19.5	559	2	S52962
43	83	19.5	559	2	S52963
44	83	19.5	559	2	S52964
45	83	19.5	559	2	S52965

ALIGNMENTS

RESULT 1
ENTRY S35221
TITLE globulin Bgl1 precursor - barley
ORGANISM Hordeum vulgare
DATE 03-Feb-1994
SEQUENCE 03-Feb-1994
TEXT CHANGE 26-Aug-1999
ACCESSIONS S35221
REFERENCE Heck, G.R.; Chamberlain, A.K.; Ho, T.H.D.
Mol. Gen. Genet. (1993) 259:209-218
Barley embryo globulin I gene, Bgl1: characterization of
cDNA, chromosome mapping and regulation of expression.
#cross-references MIM:93287988
#accession S35221
#molecule-type mRNA
#residues 1-637
#label HEC
#cross-references EMBL:M64372; NID:g167003; PID:g167004

GENETICS

#gene Bgl1
#map position 4
#classification superfamily glycyl
#keywords glycoprotein
#feature 174-190
#product globulin Bgl1
#status predicted
#label MAT
#length 637
#molecular-weight 72252
#checksum 3309

Query Match 100.0%; Score 425; DB 2; Length 637;
Best Local Similarity 100.0%; Pred. No. 2,456-60;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
ENTRY C53234
TITLE globulin-10 - maize
ORGANISM Zea mays
DATE 02-May-1994
SEQUENCE 18-Nov-1994
TEXT CHANGE 10-Oct-1997
ACCESSIONS C53234
REFERENCE Belanger, F.C.; Kriz, A.L.
Genetics (1991) 129:863-872
Molecular basis for allelic polymorphism of the maize
globulin-1 gene.

WPISEED

(TM)

Release 3.1A John F. Collins. Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sat May 13 08:42:53 2000; Maspar time 167.03 Seconds
24.906 Million cell updates/sec

Tabular output not generated.

Title: SUS-09-331-631-22
(25-847) from 450331631.pcp

Description: 459
Perfect Score: 1 EDDNHHHGHKSGCVCRRRC.....EKROERSRHEADRSGEGSS 60

Sequence: 1 EDDNHHHGHKSGCVCRRRC.....EKROERSRHEADRSGEGSS 60

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: sptrembl12
1:sp:Archea 2:sp:Bacteria 3:sp:Fungi 4:sp:Human
5:sp:Invertebrate 6:sp:Mammal 7:sp:MHC 8:sp:Organelle
9:sp:Phage 10:sp:Plant 11:sp:Rodent 12:sp:Unclassified
13:sp:Vertebrate 14:sp:Virus

Statistics: Mean 33.511; Variance 71.985; scale 0.466

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	459	100.0	122	10	GLOBULIN-1 0 ALLELE PR	9.08e-66
2	459	100.0	582	10	GLOBULIN-1 L ALLELE PR	9.08e-66
3	459	98.7	240	10	GLOBULIN-1 (FRAGMENT)	1.22e-64
4	453	98.7	540	10	GLOBULIN-1 S ALLELE PR	1.22e-64
5	450	98.0	238	10	GLOBULIN-1 (FRAGMENT)	4.50e-64
6	446	97.2	238	10	GLOBULIN-1 (FRAGMENT)	2.54e-63
7	446	97.2	238	10	GLOBULIN-1 (FRAGMENT)	2.54e-63
8	446	97.2	407	10	GLOBULIN-1 (FRAGMENT)	2.54e-63
9	445	96.9	238	10	GLOBULIN-1 (FRAGMENT)	3.92e-63
10	443	96.5	240	10	GLOBULIN-1 (FRAGMENT)	9.11e-63
11	441	96.1	242	10	GLOBULIN-1 (FRAGMENT)	2.21e-62
12	424	92.4	236	10	GLOBULIN-1 (FRAGMENT)	3.42e-59
13	424	92.4	239	10	GLOBULIN-1 (FRAGMENT)	3.42e-59
14	375	81.7	246	10	GLOBULIN-1 (FRAGMENT)	4.64e-50
15	366	79.7	242	10	GLOBULIN-1 (FRAGMENT)	4.15e-48
16	346	75.4	242	10	GLOBULIN-1 (FRAGMENT)	1.05e-44
17	343	74.7	236	10	GLOBULIN-1 (FRAGMENT)	3.75e-43
18	180	39.2	637	10	GLOBULIN-1 (FRAGMENT)	3.03e-15
19	115	25.1	393	10	EMBRIO GLOBULIN	6.82e-05
20	107	23.3	402	10	HYPOTHETICAL 45.3 KD P	1.01e-03

21	103	22.4	238	5	076510	THROMBOSPONDIN-RELATED	3.79e-03
22	99	21.6	204	13	091276	SBCFR-1 PROTEIN PRECUR	1.39e-02
23	97	21.1	444	5	044519	F56A11.6 PROTEIN.	2.65e-02
24	97	21.1	672	5	019854	F28C1.1 PROTEIN.	2.65e-02
25	96	20.9	453	2	P15320	PLASMIN PRL. COMPLETE	3.64e-02
26	94	20.5	402	4	091363	CGI-74 PROTEIN.	6.86e-02
27	92	20.0	555	10	092839	U2 SNRP AUXILIARY FAC	1.28e-01
28	91	19.8	292	10	022851	SMALL NUCLEAR RIBONUC	1.75e-01
29	91	19.8	342	11	P97485	MHC CLASS III REGION R	1.75e-01
30	91	19.8	450	4	015599	TYROSINE KINASE ACTIVA	1.75e-01
31	91	19.8	499	10	004650	LEGUMIN (FRAGMENT).	1.75e-01
32	90	19.6	155	5	021739	ROSD11.7 PROTEIN.	2.39e-01
33	90	19.6	897	5	017336	LET 858.	2.39e-01
34	90	19.6	1104	13	09M6R6	SERINE-THREONINE KINAS	2.39e-01
35	89	19.4	163	4	015410	NUCLEIC ACID BINDING P	3.25e-01
36	89	19.4	337	2	09M4V4	PUTATIVE COINTEGRATE R	3.25e-01
37	89	19.4	405	4	075939	45KDA SPLICING FACTOR.	3.25e-01
38	89	19.4	630	5	001586	COSMID K03H11.	3.25e-01
39	89	19.4	927	2	092F20	INITIATION FACTOR IF2.	3.25e-01
40	89	19.4	997	5	021721	CODED FOR BY C. ELEGAN	3.25e-01
41	89	19.4	1276	10	081059	SIMILAR TO AKONEME-ASS	4.41e-01
42	88	19.2	179	5	018409	SIMILAR TO PRE-MRNA SP	4.41e-01
43	88	19.2	515	5	045896	W06F12.2 PROTEIN	4.41e-01
44	88	15.2	573	10	092R40	U2 SNRP AUXILIARY FAC	4.41e-01
45	87	19.0	606	4	09Y341	CALCIUM CHANNEL BETA2A	5.98e-01

ALIGNMENTS

RESULT	ID	1	PRELIMINARY:	PRT:	122 AA.
AC	003863				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1999	(TREMBLrel. 12, Last annotation update)			
DE	GLOBULIN-1 0 ALLELE	PRECURSOR (GLB1-0) (7S-LIKE).			
GN	GLB1-0.				
OS	Zea mays (Maize).				
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:				
OC	euphyllophytes: Spermatophyta: Magnoliophyta: Liliopsida: Poales:				
OC	Poaceae; Zea.				
CC	[1]				
CC	SEQUENCE FROM N.A.				
CC	RN MEDLINE: 92090707.				
CC	RA BELANGER F.C., KRIZ A.L.;				
CC	*Molecular basis for allelic polymorphism of the maize globulin-1				
CC	gene.				
CC	Genetics 129:863-872(1991).				
CC	1-MISCELLANEOUS: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE				
CC	FORMATION OF THE MATURE PROTEIN FROM THE PRIMARY TRANSLATION				
CC	PRODUCT.				
CC	1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,				
CC	CONVICTILIN, CONGLYCININ, ETC.).				
CC	DR EMBL: X59085; CAA41811.1;				
CC	MEDEL: 14637; Zema:1188;14637.				
CC	KW Seed storage protein.				
CC	FT SIGNAL				
CC	FT CHAIN				
CC	FT SEQUENCE				
CC	122 AA: 13902 MW: 89C81691 CRC32:				
CC	Query Match				
CC	Best Local Similarity 100.0%: Score 459; DB 10; Length 122;				
CC	Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
CC	DB 25 EDDNHHHGHKSGCVCRRCDPRMHPRLCCEEREERKROERSRHEADRSGEGSS 84				
CC	OY 25 EDDNHHHGHKSGCVCRRCDPRMHPRLCCEEREERKROERSRHEADRSGEGSS 84				
CC	RESULT 2				
CC	ID 003865				
CC	AC 003865				
CC	DT 01-NOV-1996				
CC	PRELIMINARY:				
CC	PRT:				
CC	582 AA.				

WISNOL (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

March-pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sat May 13 08:46:12 2000; MasPar time 6.97 Seconds
406.080 Million cell updates/sec

Tabular output not generated

Title: >US-09-331-631-22

Description: (25-84) from US09221631.pep

Perfect Score: 459

Sequence: 1 EDDNNHHGHGKSGOCVRC.....EKROERSHEADRSGEGSS 60

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%

Database: Listing first 45 summaries
p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 33.532; Variance 75.335; scale 0.445

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	459	100.0	122	2	C53234	globulin-10 - maize	5.93e-05
2	459	100.0	582	2	B53234	vicillin-like storage	5.93e-63
3	453	98.7	540	2	S21825	vicillin-like storage	7.14e-62
4	453	98.7	573	2	A53234	globulin-1S, GLB1S -	7.14e-62
5	446	97.2	407	2	T02258	globulin-1 - maize	1.30e-60
6	343	74.7	236	2	T01662	globulin-1 - maize (f	2.79e-42
7	180	39.2	637	2	S35221	globulin Bgl1 precurs	1.25e-14
8	106	23.1	605	2	S06398	alpha-globulin type A	1.83e-03
9	103	22.4	1407	1	S28589	trichohyalin - rabbit	4.72e-03
10	96	20.9	325	2	JH0189	arginine/aspartate-ti	4.12e-02
11	96	20.9	382	2	S36789	gene Rb protein - hum	4.12e-02
12	96	20.9	382	2	A33640	class III histocompat	4.12e-02
13	96	20.9	588	1	FMCNAB	alpha-globulin B prec	4.12e-02
14	95	20.7	346	2	S41001	hypothetical protein	5.58e-02
15	95	20.7	1200	2	T15921	hypothetical protein	5.58e-02
16	95	20.7	1898	1	A45973	trichohyalin - human	5.58e-02
17	94	20.5	606	2	S21046	calcium channel prote	7.55e-02
18	94	20.5	632	2	S21046	calcium channel prote	7.55e-02
19	92	20.0	241	2	S44893	zx1236.7 protein - Ca	1.38e-01
20	92	20.0	466	2	T15877	hypothetical protein	1.38e-01
21	91	19.8	33	2	A41822	antimicrobial peptide	1.85e-01
22	91	19.8	375	2	A40112	MHC-region R0-repeat	1.85e-01
23	91	19.8	450	2	G01158	tyrosine kinase activ	1.85e-01

SEQ ID NO: 22

24	90	19.6	203	2	S09559	legumin A (clone pCD4	2.49e-01
25	90	19.6	507	2	T06452	probable legumin A pr	2.49e-01
26	90	19.6	517	1	FP6MLA	legumin A precursor -	2.49e-01
27	90	19.6	520	2	S08237	legumin A2 precursor	2.49e-01
28	90	19.6	722	2	S57246	ventral nervous syste	2.49e-01
29	89	19.4	163	2	I38191	nucleic acid binding	3.35e-01
30	89	19.4	407	1	EDBEQ3	immediate-early prote	3.35e-01
31	89	19.4	471	2	S02016	U1 snRNP 70K protein	3.35e-01
32	89	19.4	1276	2	T02711	hypothetical protein	3.35e-01
33	87	19.0	1335	2	S07245	xanthine dehydrogen	6.01e-01
34	86	18.7	509	2	S08059	alpha-globulin type B	8.03e-01
35	86	18.7	1380	2	T00365	hypothetical protein	8.03e-01
36	86	18.7	1473	2	T13855	suppressor of sable p	8.03e-01
37	86	18.7	2336	2	A45386	omega-conotoxin-sens1	8.03e-01
38	85	18.5	196	2	T15917	hypothetical protein	1.07e+00
39	85	18.5	197	2	A29648	female-specific trans	1.07e+00
40	85	18.5	344	2	S59043	splicing factor SRp55	1.07e+00
41	85	18.5	566	2	S22477	vicillin precursor - c	1.07e+00
42	85	18.5	668	2	A46013	coagulation factor XI	1.07e+00
43	85	18.5	915	2	B56708	extracellular signal-	1.07e+00
44	85	18.5	951	1	E64066	ribonuclease E (EC 3.	1.07e+00
45	85	18.5	2339	2	A42566	omega-conotoxin-sens1	1.07e+00

ALIGNMENTS

RESULT 1	C53234	*type complete
ENTRY	globulin-10 - maize	
TITLE	*formal_name zea mays *common_name maize	
ORGANISM	02-May-1994 #sequence-revision 18-Nov-1994 #text-change	
DATE	10-Oct-1997	
ACCESSIONS	C53234; S21823	
REFERENCE	A53234	
authors	Belanger, F.C.; Kriz, A.L.	
journal	Genetics (1991) 129:863-872	
title	Molecular basis for allelic polymorphism of the maize	
status	Globulin-1 gene.	
cross-references	MUID:92090707	
contents	Black Beauty popcorn	
accession	C53234	
status	preliminary	
molecule_type	DNA	
residues	1-122 #label BEL	
cross-references	EMBL:X59085; NID:g22281; PID:g22282	
note	Sequence extracted from NCBI backbone (NCBI:71282, NCBI:P:71286)	

SUMMARY #length 122 #molecular-weight 13902 #checksum 4000

Query Match 100.0%; Score 459; DB 2; Length 122;

Best Local Similarity 100.0%; Pred. No. 5.93e-63; Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	25	EDDNNHHGHGKSGOCVRCEDRPHORPCTLEOCREERKROERSRHEADRSGEGSS 84
Qy	25	EDDNNHHGHGKSGOCVRCEDRPHORPCTLEOCREERKROERSRHEADRSGEGSS 84

RESULT 2

ENTRY B53234 #type complete
vicillin-like storage protein G1b1-L, embryo - maize